TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Application No.	10/749,262
Filing Date	December 31, 2003
First Named Inventor	Jin-Tae Oh
Art Unit	
Examiner Name	
Total Number of Pages in This Submission: 6	Attorney Docket Number: 3364P161

ENCLOSURES ***(check all that apply)***

- [x] Fee Transmittal Form
  - [ ] Fee Attached
- [ ] Amendment / Response
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
  - [ ] PTO/SB/08
- [x] Certified Copy of Priority Document(s)
- [ ] Response to Missing Parts/ Incomplete Application
  - [ ] Basic Filing Fee
  - [ ] Declaration/POA
  - [ ] Response to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s)
- [ ] After Allowance Communication to Group
- [ ] Appeal Communication to Board of Appeals and Interferences
- [ ] Appeal Communication to Group **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [x] Other Enclosure(s) *(please identify below):*

Request for Priority; return postcard

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm or Individual name	Eric S. Hyman, Reg. No. 30,139 BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
Signature	
Date	3/16/04

CERTIFICATE OF MAILING/TRANSMISSION

I hereby certify that this correspondence is being deposited with the United States Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Typed or printed name	Melissa Stead		
Signature	Melissa Stead	Date	3-15-04

Based on PTO/SB/21 (02-04) as modified by Blakely, Sokoloff, Taylor & Zafman (wlr) 02/10/2004.
SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

FEE TRANSMITTAL for FY 2004

Effective 01/01/2004. Patent fees are subject to annual revision.

☒ Applicant claims small entity status. See 37 CFR 1.27.

TOTAL AMOUNT OF PAYMENT ($)

Complete if Known	
Application Number	10/749,262
Filing Date	December 31, 2003
First Named Inventor	Jin-Tae Oh
Examiner Name	
Art Unit	
Attorney Docket No.	3364P161

METHOD OF PAYMENT *(check all that apply)*

☐ Check ☐ Credit card ☐ Money Order ☐ Other ☐ None

☒ Deposit Account

Deposit Account Number: 02-2666

Deposit Account Name: Blakely, Sokoloff, Taylor & Zafman LLP

The Commissioner is authorized to: *(check all that apply)*

☒ Charge fee(s) indicated below ☐ Credit any overpayments

☒ Charge any additional fee(s) or underpayment of fees as required under 37 CFR §§ 1.16, 1.17, 1.18 and 1.20.

☐ Charge fee(s) indicated below, **except for the filing fee** to the above-identified deposit account

FEE CALCULATION

1. BASIC FILING FEE

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1001	770	2001	385	Utility filing fee	
1002	340	2002	170	Design filing fee	
1003	530	2003	265	Plant filing fee	
1004	770	2004	385	Reissue filing fee	
1005	160	2005	80	Provisional filing fee	

SUBTOTAL (1) ($)

2. EXTRA CLAIM FEES

	Extra Claims		Fee from below	Fee Paid
Total Claims		- 20** =		
Independent Claims		- 3 =	X	=
Multiple Dependent			X	=

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description
1202	18	2202	9	Claims in excess of 20
1201	86	2201	43	Independent claims in excess of 3
1203	290	2203	145	Multiple Dependent claim, if not paid
1204	86	2204	43	**Reissue independent claims over original patent
1205	18	2205	9	**Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) ($)

***or number previously paid, if greater, For Reissues, see below*

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1051	130	2051	65	Surcharge - late filing fee or oath	
1052	50	2052	25	Surcharge - late provisional filing fee or cover sheet.	
2053	130	2053	130	Non-English specification	
1812	2,520	1812	2,520	For filing a request for *ex parte* reexamination	
1804	920 *	1804	920 *	Requesting publication of SIR prior to Examiner action	
1805	1,840 *	1805	1,840 *	Requesting publication of SIR after Examiner action	
1251	110	2251	55	Extension for reply within first month	
1252	420	2252	210	Extension for reply within second month	
1253	950	2253	475	Extension for reply within third month	
1254	1,480	2254	740	Extension for reply within fourth month	
1255	1,210	2255	605	Extension for reply within fifth month	
1404	330	2401	165	Notice of Appeal	
1402	330	2402	165	Filing a brief in support of an appeal	
1403	290	2403	145	Request for oral hearing	
1451	1,510	2451	1,510	Petition to institute a public use proceeding	
1452	110	2452	55	Petition to revive - unavoidable	
1453	1,330	2453	665	Petition to revive - unintentional	
1501	1,330	2501	665	Utility issue fee (or reissue)	
1502	480	2502	240	Design issue fee	
1503	640	2503	320	Plant issue fee	
1460	130	2460	130	Petitions to the Commissioner	
1807	50	1807	50	Processing fee under 37 CFR 1.17(q)	
1806	180	1806	180	Submission of Information Disclosure Stmt	
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	770	1809	385	Filing a submission after final rejection (37 CFR § 1.129(a))	
1810	770	2810	385	For each additional invention to be examined (37 CFR § 1.129(b))	
1801	770	2801	385	Request for Continued Examination (RCE)	
1802	900	1802	900	Request for expedited examination of a design application	
Other fee (specify)					

* Reduced by Basic Filing Fee Paid

SUBTOTAL (3) ($)

SUBMITTED BY

Complete (if applicable)

Name (Print/Type)	Eric S. Hyman	Registration No. (Attorney/Agent)	30,139	Telephone	(310) 207-3800
Signature				Date	3/15/04

DOCKET NO.: 3364P161

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of:

JIN-TAE OH, ET AL.

Application No.: 10/749,262

Filed: December 31, 2003

For: **High-Speed Pattern Storing and Matching Method**

Art Group:

Examiner:

Commissioner for Patents
P.O, Box 1450
Alexandria, VA 22313-1450

REQUEST FOR PRIORITY

Applicant respectfully requests a convention priority for the above-captioned application, namely:

COUNTRY	APPLICATION NUMBER	DATE OF FILING
Korea	2003-0087885	5 December 2003

☒ A certified copy of the document is being submitted herewith.

Respectfully submitted,
Blakely, Sokoloff, Taylor & Zafman LLP

Dated: 3/15/04

Eric S. Hyman, Reg. No. 30,139

12400 Wilshire Boulevard, 7th Floor
Los Angeles, CA 90025
Telephone: (310) 207-3800

I hereby certify that this correspondence is being deposited with the United States Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Melissa Stead 3-15-04

Melissa Stead *Date*

대 한 민 국 특 허 청
KOREAN INTELLECTUAL PROPERTY OFFICE

별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

출 원 번 호 : 10-2003-0087885
Application Number

출 원 년 월 일 : 2003년 12월 05일
Date of Application DEC 05, 2003

출 원 인 : 한국전자통신연구원
Applicant(s) Electronics and Telecommunications Research Inst

2004 년 01 월 06 일

특 허 청
COMMISSIONER

【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0001
【제출일자】	2003.12.05
【발명의 명칭】	고속 패턴 저장 및 매칭 방법
【발명의 영문명칭】	Method of high-speed pattern storing and matching
【출원인】	
【명칭】	한국전자통신연구원
【출원인코드】	3-1998-007763-8
【대리인】	
【명칭】	유미특허법인
【대리인코드】	9-2001-100003-6
【지정된변리사】	이원일
【포괄위임등록번호】	2001-038431-4
【발명자】	
【성명의 국문표기】	오진태
【성명의 영문표기】	OH,JIN TAE
【주민등록번호】	670222-1674125
【우편번호】	305-503
【주소】	대전광역시 유성구 송강동 한솔아파트 103동 705호
【국적】	KR
【발명자】	
【성명의 국문표기】	허영준
【성명의 영문표기】	HEO,YOUNG JUN
【주민등록번호】	690106-1788519
【우편번호】	770-912
【주소】	경상북도 영천시 대창면 대창2리 644번지
【국적】	KR
【발명자】	
【성명의 국문표기】	장종수
【성명의 영문표기】	JANG,JONG SOO

【주민등록번호】 611202-1670819

【우편번호】 305-761

【주소】 대전광역시 유성구 전민동 엑스포아파트 303동 903호

【국적】 KR

【심사청구】 청구

【취지】 특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사 를 청구합니다. 대리인
유미특허법인 (인)

【수수료】

【기본출원료】	20	면	29,000 원
【가산출원료】	1	면	1,000 원
【우선권주장료】	0	건	0 원
【심사청구료】	11	항	461,000 원

【합계】 491,000 원

【감면사유】 정부출연연구기관

【감면후 수수료】 245,500 원

【기술이전】

【기술양도】 희망

【실시권 허여】 희망

【기술지도】 희망

【첨부서류】 1. 요약서·명세서(도면)_1통

【요약서】

【요약】

본 발명은 고속 패턴 저장 및 매칭 방법에 관한 것이다.

본 발명에 의하면, 일정한 규칙을 가지는 패턴 데이터를 설정된 크기에 따라 분할하고, 해당 분할된 패턴 데이터의 입력 위치순서 정보와, 해당 분할된 패턴 데이터 다음에 위치하는 패턴 데이터의 정보를 테이블화하여 저장하고, 새로 입력되는 패턴 데이터를 설정된 크기에 따라 분할하여 독립적으로 찾고, 각각의 입력 위치 순서에 따라 입력되는 패턴 데이터가 상기 일정 규칙을 가지는 패턴 데이터에 매칭 되는지 여부를 확인할 수 있도록 하여 고속으로 실시간 패턴 매칭이 가능하게 하며, 중복 단어들을 메모리의 하나의 주소에 저장할 수 있게 함으로써 메모리 효율을 높일 수 있다.

【대표도】

도 4

【색인어】

패턴 매칭, 고속 탐색, 테이블화

【명세서】

【발명의 명칭】

고속 패턴 저장 및 매칭 방법{Method of high-speed pattern storing and matching}

【도면의 간단한 설명】

도 1은 종래의 패턴 검색 방법을 위한 구조를 나타낸 블록도이다.

도 2는 종래의 패턴 검색방법을 위한 상세한 구성도이다.

도 3은 본 발명의 실시 예에 따른 IDS 규칙의 예와 단어 나누기 방법을 나타낸다.

도 4는 본 발명의 실시 예에 따른 해쉬 테이블에서의 문장 연결을 위한 구성도를 나타낸다.

【발명의 상세한 설명】

【발명의 목적】

【발명이 속하는 기술분야 및 그 분야의 종래기술】

<5> 본 발명은 고속 패턴 저장 및 매칭 방법에 관한 것으로, 특히 하드웨어로 구성된 고속 패턴 매칭 장치를 제공하여 침입 탐지 시스템을 비롯한 데이터베이스에서 특정한 패턴을 찾는 장치에 사용될 수 있도록 하는 고속 패턴 저장 및 매칭 방법에 대한 것이다.

<6> 네트워크 사용이 일반화되면서 이제까지 몇 개의 서버만이 공격의 대상이 되었던 과거의 해킹 형태를 벗어나 네트워크 전체를 무력화하고 서비스를 중단시키는 등의 네트워크 침해 대응 장치가 필요하게 되고 있다.

<7> 종래의 네트워크 기반의 침입을 탐지하기 위한 기술로는 대한민국 특허공개공보 10-2001-0012532(네트워크 기반 침입 탐지 시스템)에서 고속망에서의 네트워크 기반 침입 탐지

를 위하여 고속 하드웨어와 패턴 매칭을 하드웨어를 사용하는 네트워크 침입 탐지 엔진을 이용할 수 있도록 제안하였다.

<8> 그러나, 상기한 기술은 고속 침입 탐지에 대한 정확한 인터페이스 처리 속도와 하드웨어 콤포넌트에 대한 방법이 명시되어 있지 않는 문제가 있다.

<9> 또한, 현재까지 네트워크 침입을 찾아내기 위한 여러 방법들이 개발되고 있으며 특히 규칙 기반 패킷 매칭 방법이 가장 효과적으로 사용되고 있으며, 많은 데이터베이스에서 문장이나 단어를 찾기 위해서는 해쉬 방법을 많이 이용하고 있다.

<10> 도 1은 종래의 패턴 검색 방법을 위한 구조를 나타낸 블록도이다.

<11> 도 1을 참조하면, 패턴 검색을 위한 구성은 제어부(110), 다수의 규칙 1~n(120~140), OR 게이트(150),출력부(160) 및 레지스터(170)를 포함한다.

<12> 상기한 구조에서 제어부(110)는 각 규칙 1~n(120~140)을 제어하고, 각 규칙 1~n(120~140)은 내부의 4개의 패킷 헤드 처리기능을 위한 MAC 매칭부(121), 프로토콜부(122), IP 주소부(123), 포트 번호부(124)가 각각 MAC 주소, 프로토콜, IP 주소, 포트 번호를 정상적인 패킷의 정보와 비교하고, 각각의 비교된 결과를 AND 게이트(125)에 의해 MAC 매칭부(121), 프로토콜부(122), IP 주소부(123), 포트 번호부(124)가 모두 비교 결과 정상적임을 나타내면 AND 게이트(125)가 컨텐츠 패턴 매치부(126)가 정상적인 패킷임을 나타내는 신호를 출력하도록 하는 신호를 인가한다.

<13> 그리고, 모든 규칙 1~n(120~140)의 컨텐츠 패턴 매칭부(126)의 신호를 OR 연산하는 OR 게이트(150)를 통해 하나의 규칙 1~n(120~140)라도 정상적인 패킷이 아니라는 신호를 나타내면

패킷 출력부(160)가 에러 신호를 발생하도록 하고, 모든 규칙 1~n(120~140)이 정상적인 패킷임을 나타내면 패킷 출력부(160)가 해당 패킷을 출력하도록 한다.

<14> 상기에서 모든 규칙 1~n(120~140) 부분은 FPGA(Field Programmable Gate Array)에 프로그램되어 구성되며, 규칙의 갯수에 따라 프로그램이 달라지게 된다.

<15> 상기의 패킷 검색을 좀더 자세히 설명하면 다음과 같다.

<16> 도 2는 종래의 패턴 검색방법을 위한 상세한 구성도이다.

<17> 도 2를 참조하면, 32비트 레지스터(127)로 입력되는 문자열의 패턴을 검색하는 컨텐츠 패턴 매치부(126)는 32비트 단위로 입력되는 데이터에서 'patterns' 이라는 문자열이 3클럭 동안 전달되는 경우를 예를 들면, 처음 32비트에는 Cyc(Cycle) 1에서 보인 'pat'라는 문자열이 포함되어 있고, Cyc 2에는 다음 클럭에 전달되어 온 'tern'이라는 문자열이, 그리고 Cyc 3에서는 's'를 포함하고 있다.

<18> 즉, col 1에는 문자열 'patterns'라는 단어가 row 1의 처음 바이트부터 비교하는 것을 보이고 있다. 즉, col 1에서는 row 1에서 4바이트 'patt'를 비교하고 raw 2에서는 'erns'를 동시에 비교하여 결과를 내놓게 된다. 이 경우 레지스터 A에 있는 문자열은 처음 바이트부터 다른 문자열이므로 비교값이 거짓이 된다.

<19> 또한, col 2에서는 col 1에서 비교하던 문자열은 한 바이트씩 아래로 시프트 한 값으로 입력 값을 비교한다. 즉, col 2의 row 1에서는 첫 바이트는 무시하고, 두 번째 바이트부터 'pat' 3바이트를 비교하고, row 2에서는 'tern'을 비교하고 row 3에서는 's'만을 비교하면 입력된 문자열은 col 2에서 비교 결과가 참이 된다.

<20> 동일한 방법으로 col 3과 col 4에서도 문자열을 한 바이트식 아래로 시프트하여 비교한다. 따라서 col 1, col 2, col 3, col 4의 비교값을 OR(129) 논리 합으로 구한 것이 매치 신호로 발생한다.

<21> 그러나, 이러한 방법은 하드웨어를 기반으로 패턴 매칭 장치를 설계한 것으로, 규칙의 수가 증가할 경우 FPGA를 새로 프로그램 해야 하며, 많은 규칙에 대해서 회로의 복잡도고 증가하므로 원하는 속도를 얻는 것에 문제가 있다.

【발명이 이루고자 하는 기술적 과제】

<22> 상기한 문제를 해결하기 위하여, 본 발명은 고속 패턴 매칭을 위하여 모든 구성이 단순한 메모리 룩업으로 구성될 수 있도록 하고, 새로운 룰의 추가 및 갱신이 용이하게 하여, 새로운 패턴을 계속 추가해야 하는 장치에도 적용이 용이하도록 하고, 규칙 기반 IDS(Intrusion Detection System)의 패턴 매칭용 하드웨어에 적용될 수 있으며, 특정 패턴을 빠른 시간에 찾아야 하는 분야에 적용될 수 있도록 하는 고속 패턴 저장 및 매칭 방법을 제공함에 그 목적이 있다.

【발명의 구성 및 작용】

<23> 본 발명의 하나의 특징에 따른 고속 패턴 저장 방법은,

<24> 규칙을 구성하는 패턴 데이터를 테이블화하여 저장하는 방법에 있어서, (a) 상기 패턴 데이터를 설정된 길이로 분할하는 단계; (b) 상기 분할된 각각의 패턴 데이터의 입력 위치순서 정보를 추출하는 단계; 및 (c) 상기 분할된 각각의 패턴 데이터에 고유의 패킷 ID를 부여하고, 분할된 패턴 데이터 및 해당 패턴 데이터의 입력 위치순서 정보를 테이블화하여 저장하는 단계를 포함한다.

<25> 본 발명의 하나의 특징에 따른 고속 패턴 매칭 방법은,

<26> 패턴 데이터를 입력 순서 정보를 포함한 특정 규칙에 의해 테이블화하여 저장하고, 입력되는 데이터의 패턴과 상기 패턴 데이터와 매칭 되는지 여부를 판단하는 고속 패턴 매칭 방법에 있어서, (a) 상기 입력되는 데이터를 상기 특정 규칙에 따라, 일정 크기로 분할하는 단계; (b) 상기 분할된 데이터와 동일한 데이터를 저장한 테이블 정보를 검색하는 단계; (c) 상기 검색된 각각의 분할 데이터와 동일한 데이터를 저장한 테이블 정보에 포함된 해당 데이터의 입력위치 순서정보와, 상기 분할된 데이터의 입력위치 순서정보가 같은 테이블 정보를 추출하는 단계; 및 (d) 상기 추출된 테이블 정보에 의하여 상기 입력되는 데이터와 같은 패턴의 데이터가 구성되는지를 판단하는 단계를 포함한다.

<27> 아래에서는 첨부한 도면을 참고로 하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시 예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다. 첨부된 도면은 본 발명을 명확하게 설명하기 위해 본 발명의 설명과 관계없는 부분은 생략하였으며, 동일 또는 유사한 부분에 대해서는 동일한 도면 부호를 붙였다.

<28> 본 발명의 실시 예에 따라 침입 탐지 규칙을 찾는데 있어서, 먼저 침입 탐지 규칙을 구성하는 문장은 동일한 위치에서 동일한 문자열들을 가지는 규칙이 많이 존재한다.

<29> 따라서 규칙을 구성하는 문장을 앞에서부터 일정한 길이 이하로 잘라 이를 단어로 정의하고 각 단어를 따로 테이블에서 찾고 이들을 연결하여 규칙을 찾을 수 있다.

<30> 또한, 단어를 나눌 때는 위치가 다른 곳에서 동일한 단어가 반복되는 일은 거의 없거나 혹은 반복되는 단어가 있을 경우 해당 규칙의 문제가 발생한 위치에서만 단어를 자르는 길이를

달리하면 단어가 문장의 다른 위치에서 반복되는 것을 막을 수 있으므로, 각 단어가 지닌 순서 정보의 결합이 서로 독립적인 특성을 이용하여 위치에 따라 비교해야 할 패턴의 수가 항상 규칙의 수보다 적거나 동일하고, 각 단어를 별개로 찾아내고 단어들의 순서를 적절히 연결하여 정확한 규칙을 찾을 수 있다.

<31> 도 3은 본 발명의 실시 예에 따른 IDS 규칙의 예와 단어 나누기 방법을 나타낸다.

<32> 도 3을 참조하면, 본 발명의 실시 예에 따라 오픈 소스 IDS인 snort의 웹-공격 (Web-attack) 규칙들 중에서 8개의 규칙을 예를 들어 설명한다.

<33> 이때, 8개의 규칙은 도 3의 제 1 블록(310)에 나타낸 것이고, 해쉬 테이블을 구성하기 위해 규칙들 중 반복되는 문장들을 추출하여 최대 길이 7 바이트 이내로 자르고 이들의 연결 관계를 보이는 것이 제 2 블록(320)에 나타낸다.

<34> 상기한 도 3과 같은 실시 예를 이용하여 컴퓨터에서 "/bin/echo"를 찾고자 하는 경우 먼저 "/bin/"이라는 단어를 찾는 것을 설명하면 다음과 같다.

<35> 종래에는 먼저"/bin/"이라는 단어를 찾은 다음에는 다음에 올 수 있는 3 가지의 단어들을 지시하는 포인터를 찾아 "echo", "kill", 그리고 "chomod" 중의 하나와 데이터를 차례로 비교하는 방법을 사용한다.

<36> 이러한 방법은 "/bin/"을 찾은 다음, 다음 3개의 문장과 입력되는 데이터를 차례로 비교해야 하므로 데이터 비교 시간이 비교할 데이터양만큼 증가하게 된다.

<37> 이때, 본 발명의 실시 예에 따른 도 3에 나타난 바와 같이 제 2 블록(320)의 데이터 구조를 이용하여 데이터를 저장하면 데이터 저장 공간을 줄일 수 있는 장점이 있다.

<38> 그러나, 이때 실시간으로 입력되는 패킷 등에서 원하는 패턴을 찾기 위한 실시간성 보장의 문제가 발생하므로, 본 발명의 실시 예에 따라 도 3의 제 2 블록의 데이터를 다중 해쉬 테이블에 각 단어의 해쉬 값에 따라 저장하게 된다.

<39> 상기한 방법에 의하면 입력되는 문장에서 잘려진 단어들을 해쉬 테이블에서 각각 찾아내고 각 단어가 저장되어 있던 위치 정보와 함께 출력할 수 있다.

<40> 해쉬 테이블에서 검색된 각 단어와 각 단어의 해쉬 테이블에 저장되어 있던 위치정보를 이용하면 각 단어의 전후를 비교하여 찾고자 하는 전체 문장을 찾을 수 있다.

<41> 다음은 해쉬 테이블에 각 단어가 저장될 때의 연결 관계를 나타낸 것이다.

<42> 도 4는 본 발명의 실시 예에 따른 해쉬 테이블에서의 문장 연결을 위한 구성도를 나타낸다.

<43> 도 4를 참조하면, 각 단어가 해쉬 테이블에 저장될 때 이들의 연결 관계를 보이기 위해 해쉬 테이블의 각 주소에는 해당 단어의 이전 ID(pid; Previous-ID)와 해당 단어의 ID(mid)에 대한 데이터를 가지고 있다.

<44> 여기서 해당 단어의 ID는 해당 단어가 저장된 메모리 주소를 대신하여 사용할 수도 있다.

<45> 본 발명의 실시 예에 따른 도 4에서 각 단어들은 다중의 해쉬 테이블을 사용하여 저장될 수 있으며, 제 1 테이블(410)은 "/bin/"이 저장된 해쉬 테이블의 주소를 나타낸다.

<46> 제 1 테이블(410)은 이 주소에 해당하는 단어 이전의 어떠한 단어가 있었는지를 pid1 로 연결하는 것이 보이며, 제 1 테이블(410)에 나타난 "/bin/"이 규칙을 구성하는 처음 단어이며, 이러한 첫 단어의 경우에는 pid 1에 다른 정보를 저장할 수 있다.

<47> 예를 들어, 처음 단어를 저장하는 제 1 테이블(410)과 제 2 테이블(420) 및 제 3 테이블(430)에 저장된 단어는 문장의 첫 단어이므로, pid1, pid2, pid3에 이전 단어의 정보가 아닌 본 발명의 실시 예에 따라 HTTP 프로토콜을 사용하는 규칙에 의한 HTTP ID 저장하여, 예시된 규칙들이 HTTP 프로토콜에서만 찾아질 수 있도록 한다. 그리고, 각 Ctl 1, Ctl2, Ctl3, 에 해당 단어가 첫번째 단어임을 나타내는 정보로'1'이라는 숫자를 부여하고, 두 번째 단어를 저장하고 있는 제 4 테이블(440), 제 5 테이블(450), 제 6 테이블(460) 및 제 7 테이블(470)의 Ctl4, Ctl5, Ctl6, Ctl7에 '2'라는 숫자를 부여하여 각 단어가 정확한 위치에서 찾아지고 조합되는지를 확인하는 수단으로 사용할 수 있다.

<48> 또한, "echo"의 단어와 같이 두 번째 단어이면서 마지막 단어인 경우는 Ctl4에 두 번째 단어임을 나타내는 '2'의 정보와 함께 해당 단어가 마지막 단어라는 정보도 포함하여 저장함으로써, 마지막 단어 정보가 있는 단어를 검색한 이후에는 더이상 검색을 하지 않도록 한다.

<49> 상기한 도 4에서 입력되는 패킷이 HTTP 프로토콜을 사용하고 "/bin/echo"라는 문장을 포함하는 경우 각각의 단어 "/bin/"과 "echo"를 도 4의 제 1 테이블(410)과, 제 4 테이블(440)에서 찾을 수 있다.

<50> 이때, 제 1 테이블에서 HTTP 프로토콜을 위한 ID를 pid에 저장하고 있다면, 읽혀진 pid1과 패킷의 헤드에서 HTTP 프로토콜임이 확인되어 발생한 ID를 비교하고, HTTP 프로토콜을 사용한 패킷의 "/bin/"이 있음을 확인하면 문장을 계속 찾아갈 수 있다.

<51> 이때 만약 입력되는 패킷이 HTTP 프로토콜을 사용하지 않는다면 프로토콜 비교에 있어 거짓이 발생하므로 처음 단어 "/bin/"는 맞게 찾는 단어가 아니게 되어 룩업의 결과는 거짓이 된다.

<52> 또한, 다음 단어인 "echo"를 저장하고 있는 제 4 테이블(440)의 pid 4는 mid 1과 연결되어 있으므로, 두 단어가 연결되어 있음을 확인하고 제 1 테이블(410)의 pid 1에는 해당 단어가 첫 번째 단어라는 정보를 포함하며, 제 4 테이블(440)의 pid 4에는 해당 단어가 두 번째 단어이면서 마지막 단어라는 정보를 포함하고 있으므로 완성된 문장을 찾을 수 있다.

<53> 또한, 단어 사이의 간격정보를 사용하여 문장 중에 있는 메타문자-예를 들어 mat*.dat-를 처리할 수 있다. 즉, 예를 든 바와 같이 'mat*.dat'가 찾고자 하는 문장인 경우, 'mat'와 'dat'를 단어로 각각 찾고, 두 단어 사이의 다른 단어나 문자가 들어갈 수 있다는 정보를 간격 정보로 각 단어가 저장된 테이블에 저장한다.

<54> 그리고, 각 단어의 연결을 확인할 때 이 간격 정보를 사용하여 메타문자를 처리한다. 이러한 정보들은 각 테이블에서 Ctl 필드를 이용한다. 이러한 메타문자의 처리는 패턴 검색에서 꼭 필요하는 기능이며, 하드웨어에서 처리하기가 쉽지 않은 기능이다.

<55> 이상과 같은 본 발명의 실시 예에서와 같이 제 1 테이블~제 7 테이블로 해쉬 테이블을 작게 다중으로 사용하는 방법을 사용하여 동시에 같은 해쉬 값을 가진 다른 단어를 찾을 수 있도록 하였다.

<56> 또한, 이러한 경우의 단어를 해쉬 테이블에서 입력된 단어와 정확히 맞고 단어가 나타날 위치가 정확한가를 검사하는 과정이 있어, 검색을 통해 찾아낸 결과가 유일하게 정의 될 수 있도록 한다.

<57> 본 발명의 실시 예와 같이 단어의 검색에 해쉬 테이블을 사용하는 경우, 해쉬 키의 충돌을 막기 위해서 작은 해쉬 테이블을 다중으로 사용하는 방법으로 동시에 같은 해쉬 값을 가진 다른 단어들을 찾을 수 있도록 하였다.

<58> 또한, 다중 테이블을 이용하는 방법이 원하는 단어를 정확히 찾아낼 수 없는 문제를 해결하기 위해서, 단어를 해쉬 테이블에 입력된 단어와 정확히 맞고, 단어가 나타날 위치가 정확히 맞는지 검사하는 과정을 거쳐 유일한 단어가 정의될 수 있도록 한다.

<59> 그리고, 입력되는 테이블에서 발생하는 해쉬 값에 따라 각 테이블을 매번 읽는 것은 하드웨어 적으로 많은 전력 소모를 가져오게 되므로, 단어들이 문장에서 나타날 순서 정보를 테이블에 따로 저장하고, 먼저 이 순서 정보를 읽어 단어의 순서가 맞는지를 확인하여 단어를 비교하도록 하여 단어를 비교하기 위해 테이블을 읽는 빈도를 줄일 수 있도록 한다.

<60> 필요한 경우, 하나의 공통된 단어를 접미사로 사용하는 방법으로 접미사로 선택된 단어를 제외한 문장까지를 하나의 문장으로 놓고 단어를 구분하여 마지막에 end를 넣고, 해당 단어의 하나의 D를 부여함으로써, 같은 접미사를 갖는 문장의 마지막 단어들이 동일한 ID를 지니도록 하여 동일한 접미사를 가진 문장 처리가 용이하록 한다.

<61> 또한, 작은 크기의 해쉬 테이블은 해쉬 비트 수가 적어서, 단어의 길이가 동일하고 서로 독립인 단어가 동일한 해쉬 값을 생성하는 경우가 많다. 따라서 해쉬 값을 이용하여 해쉬 테이블을 찾은 다음 입력된 문자열과 테이블에 저장되어 문자열을 직접 비교하여 입력된 문자열이 찾고자 하는 문자인지 비교하는 기능을 필요로 한다.

<62> 따라서 해쉬 테이블에 저장된 단어의 길이가 짧으면 하드웨어 구현이 쉽고 이를 고려하여 본 발명의 실시 예에서는 해쉬 테이블에 저장될 단어를 특정한 길이 이하로 잘라 저장하고, 이를 비교하도록 하는 방법을 제안한다.

<63> 이상에서 본 발명의 바람직한 실시 예에 대하여 상세하게 설명하였지만 본 발명은 이에 한정되는 것은 아니며, 그 외에 다양한 변경이나 변형이 가능하다.

【발명의 효과】

<64> 이상에서 설명한 바와 같이, 본 발명에 따른 고속 패턴 저장 및 매칭 방법은 단순한 메모리 룩업으로 구성되어 새로운 룰의 추가와 갱신이 용이하여 새로운 패턴을 계속해서 추가하여 검색하는데 용이하며, 규칙 기반의 IDS 또는 지문 비교, DNS 비교 등 많은 데이터 중에서 빠른 시간 안에 특정 패턴을 찾아야 하는 분야에 적용하여 패턴 매칭이 고속으로 실현될 수 있도록 하는 효과가 있다.

<65> 또한, 각 단어가 지닌 순서 정보의 결합이 서로 독립적인 특성을 이용하여 단어들을 단어 정보가 저장된 테이블에서 찾고 하나의 단어를 찾을 때 앞 단어의 ID를 비교하는 방법으로 문장을 연결하여 실시간 패턴 검색이 가능하도록 하는 효과가 있다.

【특허청구범위】

【청구항 1】

규칙을 구성하는 패턴 데이터를 테이블화하여 저장하는 방법에 있어서,

(a) 상기 패턴 데이터를 설정된 길이 이하로 분할하는 단계;

(b) 상기 분할된 각각의 패턴 데이터의 입력 위치순서 정보를 추출하는 단계; 및

(c) 상기 분할된 각각의 패턴 데이터에 고유의 패킷 ID를 부여하고, 분할된 패턴 데이터 및 해당 패턴 데이터의 입력 위치순서 정보를 테이블화하여 저장하는 단계

를 포함하는 것을 특징으로 하는 고속 패턴 저장 방법.

【청구항 2】

제 1항에 있어서,

상기 테이블 정보는,

해당 테이블 정보에 저장된 패턴 데이터의 다음 순서의 입력 위치를 가지는 패턴 데이터의 고유 패턴 ID 정보를 포함하는 것을 특징으로 하는 고속 패턴 저장 방법.

【청구항 3】

제 1항에 있어서,

해당 패턴 데이터의 간격정보를 포함하여 메타문자 처리를 하도록 하는 것을 특징으로 하는 고속 패턴 저장 방법.

【청구항 4】

제 1항에 있어서,

상기 (c) 단계에서,

상기 분할된 패턴 데이터 중에서 입력 위치순서가 처음인 패턴 데이터의 경우, 패킷 헤드의 정보를 고유의 패킷 ID로 설정하는 것을 특징으로 하는 고속 패턴 저장 방법.

【청구항 5】

제 1항에 있어서,

상기 (c)단계에서,

해당 테이블에 저장되는 패턴 데이터 또는 해당 패턴 데이터의 입력위치 순서 또는 해당 패턴 데이터의 다음에 오는 패턴 데이터가 다른 경우, 별도의 테이블에 저장하여 다중화하는 것을 특징으로 하는 고속 패턴 저장 방법.

【청구항 6】

제 1항에 있어서,

상기 가장 마지막 순서에 해당하는 분할된 패턴 데이터가 동일한 여러 패턴 데이터에 대하여 해당 마지막 순서의 분할된 패턴 데이터가 같은 위치 정보를 갖도록 저장하는 것을 특징으로 하는 고속 패턴 저장 방법.

【청구항 7】

제 1항에 있어서,

상기 (c)단계에서,

상기 분할된 패턴 데이터가 가장 마지막 위치에 있는 경우, 해당 패턴 데이터가 가장 마지막 순서의 패턴 데이터임을 나타내는 정보를 입력위치 순서 정보에 포함시키는 것을 특징으로 하는 고속 패턴 저장 방법.

【청구항 8】

제 1항에 있어서,

상기 패턴 데이터는 해쉬 테이블에 저장하도록 하고, 각각의 분할된 패턴 데이터의 해쉬 값, 해당 분할된 패턴 데이터의 순서정보, 단어의 연결관계 정보를 저장하는 것을 특징으로 하는 고속 패턴 저장 방법.

【청구항 9】

입력되는 데이터의 패턴과 특정 규칙에 의해 테이블화하여 저장된 패턴 데이터의 매칭 여부를 판단하는 고속 패턴 매칭 방법에 있어서,

(a) 상기 입력되는 데이터 패턴을 설정된 크기로 이하로 분할하는 단계;

(b) 상기 분할된 데이터 패턴과 동일한 패턴 데이터를 저장한 테이블 정보를 검색하는 단계;

(c) 상기 검색된 각각의 분할 데이터 패턴과 동일한 패턴 데이터를 저장한 테이블 정보에 포함된 해당 데이터의 입력위치 순서정보와, 상기 분할된 데이터 패턴의 입력위치 순서정보가 같은 테이블 정보를 추출하는 단계; 및

(d) 상기 추출된 테이블 정보에 의하여 상기 입력되는 데이터 패턴과 같은 패턴 데이터가 구성되는지를 판단하는 단계

를 포함하는 고속 패턴 매칭 방법.

【청구항 10】

제 9항에 있어서,

상기 패턴 데이터는,

해당 패턴 데이터의 입력위치 순서를 나타내는 패킷 ID, 및

해당 패턴 데이터의 입력 위치 다음에 오는 패턴 데이터의 패킷 ID

정보를 포함하여 저장되는 것을 특징으로 하는 고속 패턴 매칭 방법.

【청구항 11】

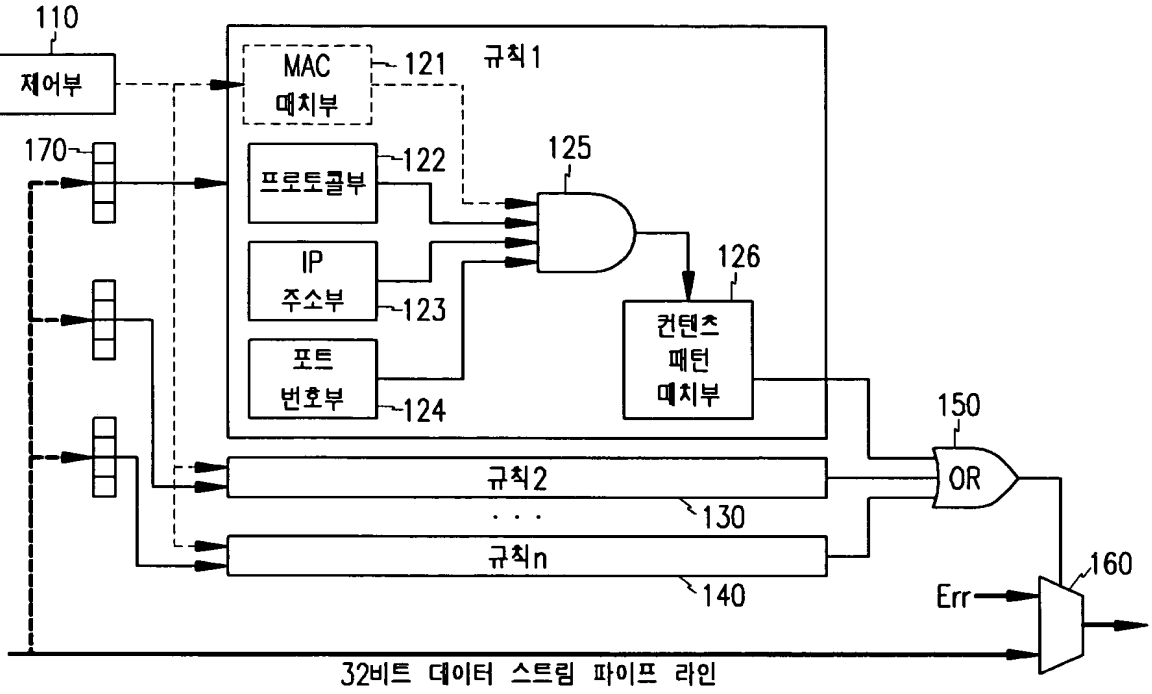
제 9항에 있어서,

상기 (b) 단계에서,

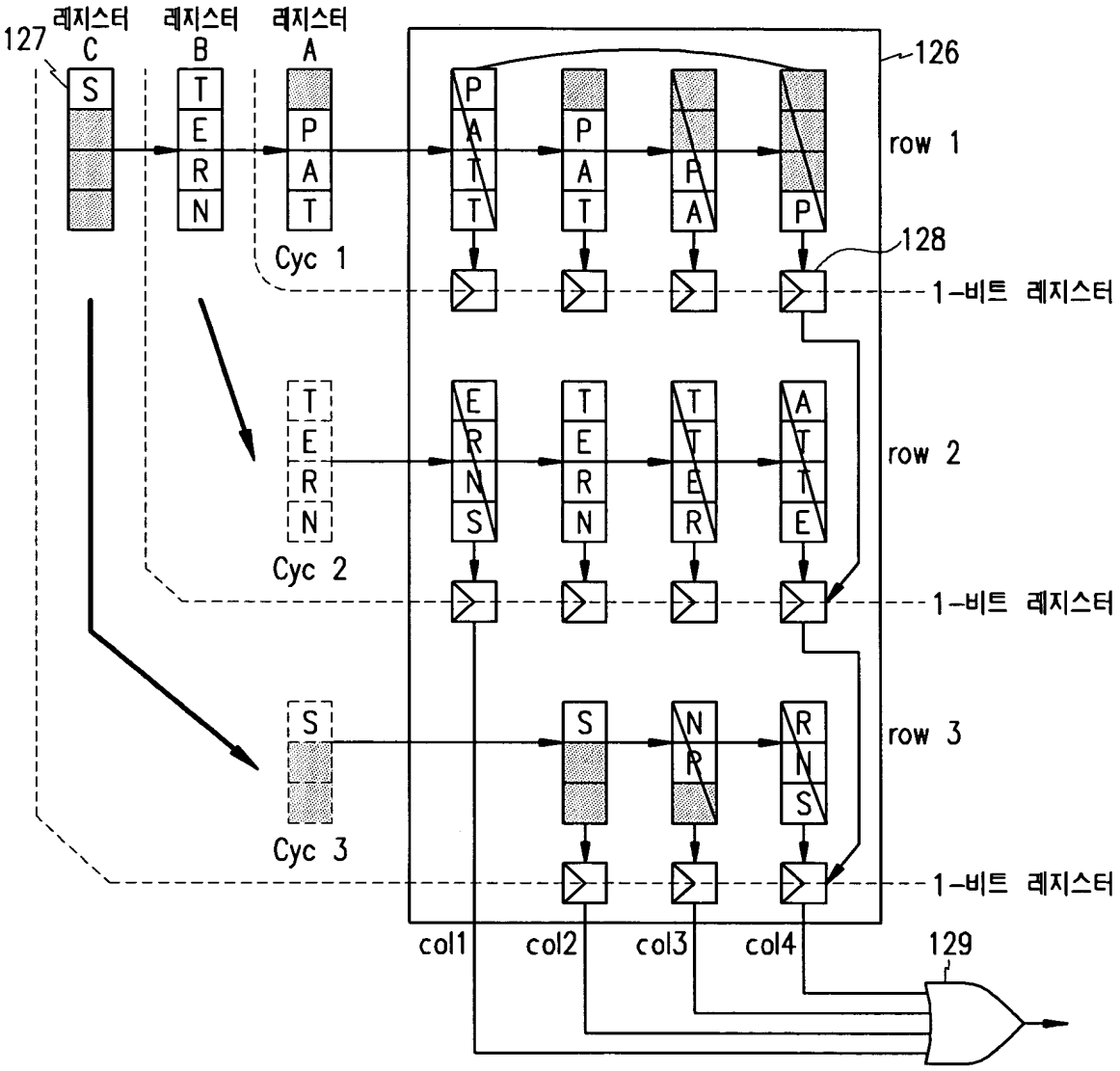
상기 분할된 데이터 패턴의 입력위치 정보와, 해당 데이터 패턴과 동일한 패턴 데이터의 입력위치저보와 틀린 경우, 해당 패턴 데이터와 연결되는 패턴 데이터에 대한 검색을 중지하는 단계를 더 포함하는 고속 패턴 매칭 방법.

【도면】

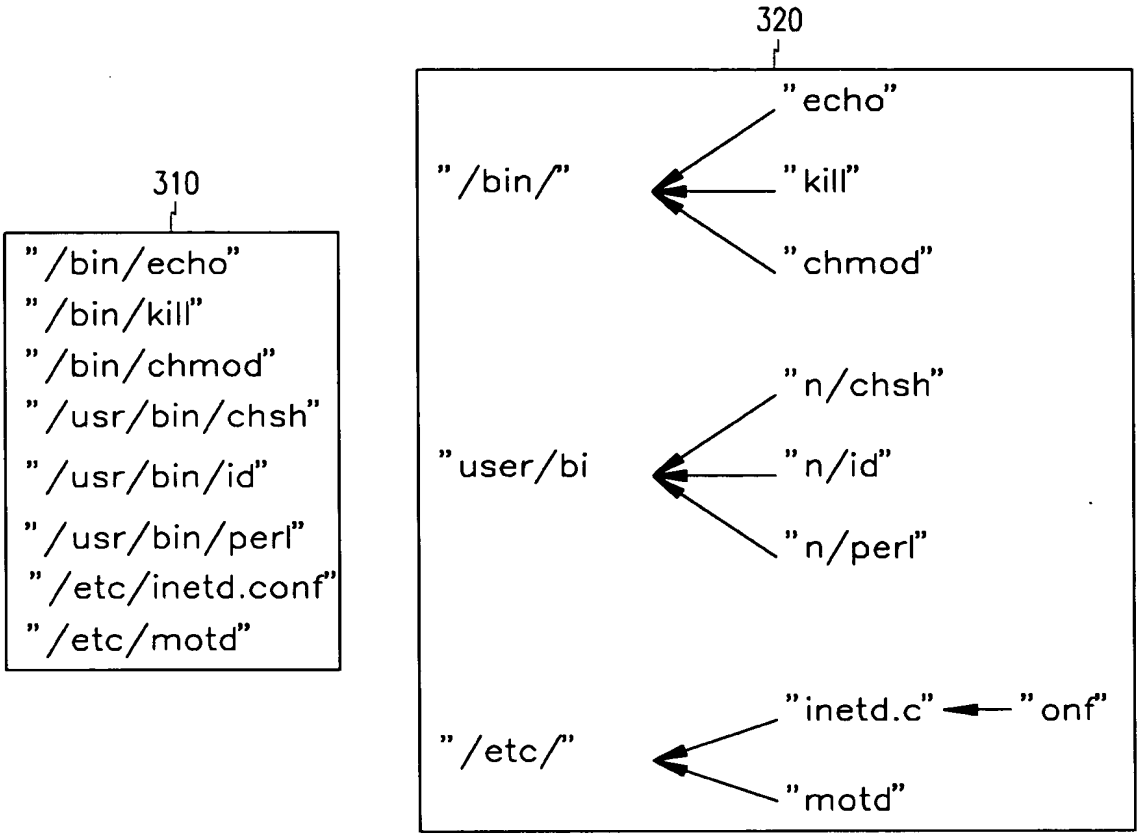
【도 1】


110
제어부
MAC
매치부
121
규칙1
170
프로토콜부
122
125
IP
주소부
123
126
포트
번호부
124
컨텐츠
패턴
매치부
150
OR
규칙2
130
규칙n
140
160
Err
32비트 데이터 스트림 파이프 라인

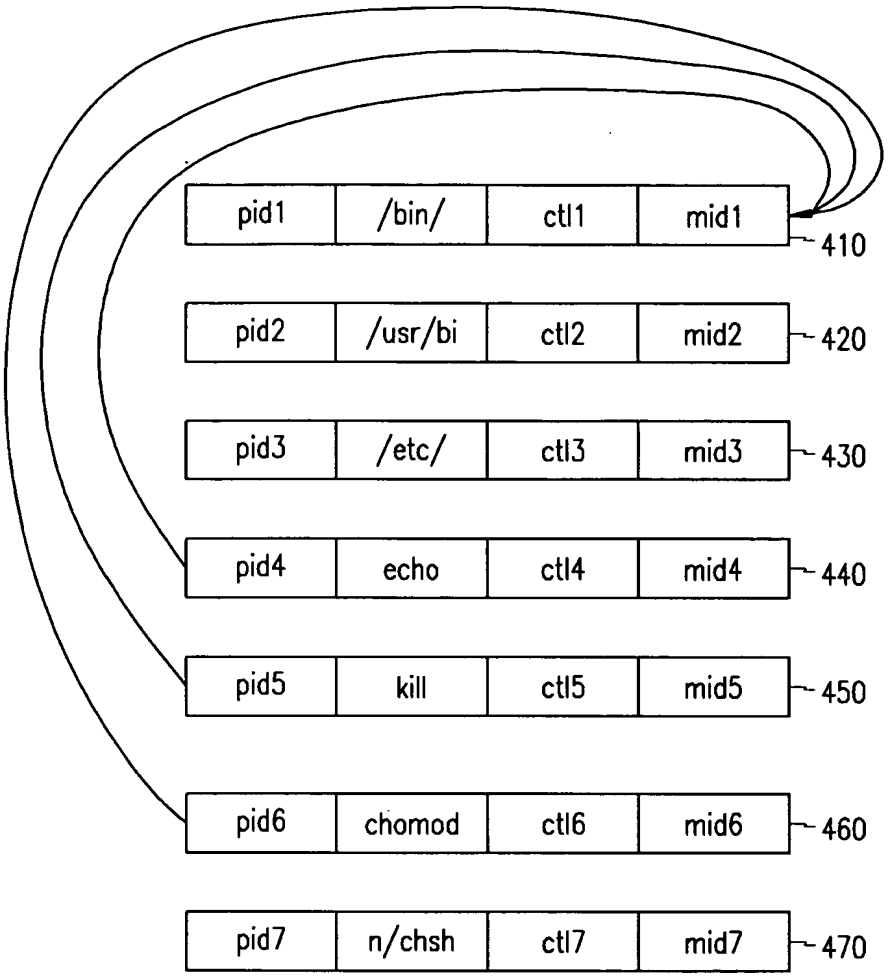
【도 2】


127
레지스터 C
레지스터 B
레지스터 A
S
T E R N
P A T
Cyc 1
T E R N
Cyc 2
S
Cyc 3
126
row 1
row 2
row 3
128
1-비트 레지스터
1-비트 레지스터
1-비트 레지스터
col1
col2
col3
col4
129

【도 3】


310
"/bin/echo"
"/bin/kill"
"/bin/chmod"
"/usr/bin/chsh"
"/usr/bin/id"
"/usr/bin/perl"
"/etc/inetd.conf"
"/etc/motd"
320
"/bin/"
"echo"
"kill"
"chmod"
"user/bi
"n/chsh"
"n/id"
"n/perl"
"/etc/"
"inetd.c"
"onf"
"motd"

【도 4】


pid1
/bin/
ctl1
mid1
410
pid2
/usr/bi
ctl2
mid2
420
pid3
/etc/
ctl3
mid3
430
pid4
echo
ctl4
mid4
440
pid5
kill
ctl5
mid5
450
pid6
chomod
ctl6
mid6
460
pid7
n/chsh
ctl7
mid7
470